

# INVERSE PROBLEMS FOR LINEAR FORMS OVER FINITE SETS OF INTEGERS

MELVYN B. NATHANSON

ABSTRACT. Let  $f(x_1, x_2, \dots, x_m) = u_1x_1 + u_2x_2 + \dots + u_mx_m$  be a linear form with positive integer coefficients, and let  $N_f(k) = \min\{|f(A)| : A \subseteq \mathbf{Z} \text{ and } |A| = k\}$ . A minimizing  $k$ -set for  $f$  is a set  $A$  such that  $|A| = k$  and  $|f(A)| = N_f(k)$ . A finite sequence  $(u_1, u_2, \dots, u_m)$  of positive integers is called complete if  $\left\{\sum_{j \in J} u_j : J \subseteq \{1, 2, \dots, m\}\right\} = \{0, 1, 2, \dots, U\}$ , where  $U = \sum_{j=1}^m u_j$ . It is proved that if  $f$  is an  $m$ -ary linear form whose coefficient sequence  $(u_1, \dots, u_m)$  is complete, then  $N_f(k) = Uk - U + 1$  and the minimizing  $k$ -sets are precisely the arithmetic progressions of length  $k$ . Other extremal results on linear forms over finite sets of integers are obtained.

## 1. EXTREMAL FUNCTIONS FOR LINEAR FORMS

Let  $m \geq 1$  and let  $f : \mathbf{Z}^m \rightarrow \mathbf{R}$  be a real-valued function of  $m$  integer variables. For every finite set  $A$  of integers, consider the set

$$f(A) = \{f(a_1, a_2, \dots, a_m) : a_1, a_2, \dots, a_m \in A\}.$$

Let  $|A|$  denote the cardinality of the set  $A$ . We define the functions

$$N_f(k) = \min\{|f(A)| : A \subseteq \mathbf{Z} \text{ and } |A| = k\}$$

and

$$M_f(k) = \max\{|f(A)| : A \subseteq \mathbf{Z} \text{ and } |A| = k\}.$$

A set  $A$  with  $|A| = k$  is called a  $k$ -set. If  $A$  is a  $k$ -set and  $|f(A)| = N_f(k)$ , then  $A$  is called a *minimizing  $k$ -set* for  $f$ . If  $A$  is a  $k$ -set and  $|f(A)| = M_f(k)$ , then  $A$  is called a *maximizing  $k$ -set* for  $f$ . An important inverse problem in number theory is to compute the extremal functions  $N_f(k)$  and  $M_f(k)$ , and to classify the minimizing and maximizing  $k$ -sets for  $f$ .

In this paper we study linear forms. A classical example in additive number theory is the linear form  $f(x_1, x_2, \dots, x_m) = x_1 + x_2 + \dots + x_m$ . In this case,  $N_f(k) = mk - m + 1$  and the minimizing  $k$ -sets are the arithmetic progressions of length  $k$  (Nathanson [3, Theorem 1.6]). Also,  $M_f(k) = \binom{k+m-1}{m} = k^m/m! + O(k^{m-1})$  and the maximizing  $k$ -sets are sets  $A$  of positive integers (called Sidon sets or  $B_h$ -sets) such that every integer has at most one representation as the sum of  $h$  not necessarily distinct elements of  $A$ .

---

*Date:* February 1, 2008.

*2000 Mathematics Subject Classification.* Primary 11P70, 11B25, 11B37, 11B75, 11A25.

*Key words and phrases.* Inverse problem, arithmetic progression, linear form, sumset, extremal problem, additive number theory, combinatorial number theory.

This work was supported in part by grants from the NSA Mathematical Sciences Program and the PSC-CUNY Research Award Program.

Let  $\mathcal{LF}(m)$  denote the set of all  $m$ -ary linear forms

$$f(x_1, \dots, x_m) = u_1x_1 + u_2x_2 + \dots + u_mx_m$$

with positive integer coefficients. The function  $f$  is called the *linear form associated to the sequence*  $(u_1, \dots, u_m)$ . Without loss of generality we can assume that

$$1 \leq u_1 \leq u_2 \leq \dots \leq u_m$$

and

$$\gcd(u_1, u_2, \dots, u_m) = 1.$$

Let  $\mathcal{LF}^*(m)$  denote the set of all  $m$ -ary linear forms with pairwise distinct coefficients, that is, with

$$1 \leq u_1 < u_2 < \dots < u_m.$$

We define the extremal functions

$$\mathcal{N}_m(k) = \min\{N_f(k) : f \in \mathcal{LF}(m)\}$$

and

$$\mathcal{N}_m^*(k) = \min\{N_f(k) : f \in \mathcal{LF}^*(m)\}.$$

Since the only linear form with  $m = 1$  is  $f(x_1) = u_1x_1$ , it follows that  $\mathcal{N}_1(k) = k$  for all  $k \geq 1$ . For binary forms we have  $\mathcal{N}_2(2) = 3$  and  $\mathcal{N}_2^*(2) = 4$ .

Fix the integer  $m \geq 1$ . For  $k \geq 2$  and  $f \in \mathcal{LF}(m)$ , let  $A$  be a  $k$ -set such that  $N_f(k) = |f(A)|$  and let  $a' = \max(A)$  and  $A' = A \setminus \{a'\}$ . Since  $f(A) \supseteq f(A')$  and  $f(a', \dots, a') > \max(f(A'))$ , it follows that

$$N_f(k) = |f(A)| > |f(A')| \geq N_f(k-1)$$

and so  $\mathcal{N}_m(k) > \mathcal{N}_m(k-1)$  and  $\mathcal{N}_m^*(k) > \mathcal{N}_m^*(k-1)$  for all  $k \geq 2$ . Choosing a  $(k-1)$ -set  $A'$  such that  $M_f(k-1) = |f(A')|$  and an integer  $a' > \max(A')$ , we define the  $k$ -set  $A = A' \cup \{a'\}$ . Since  $f(A) \supseteq f(A')$  and  $f(a', \dots, a') > \max(f(A'))$ , we have

$$M_f(k-1) = |f(A')| < |f(A)| \leq M_f(k).$$

Similarly, for fixed  $k \geq 2$ , the extremal functions  $\mathcal{N}_m(k)$  and  $\mathcal{N}_m^*(k)$  are strictly increasing in  $m$ .

Denote the interval of integers  $\{n \in \mathbf{Z} : x \leq n \leq y\}$  by  $[x, y]$ . Given a finite sequence of integers  $\mathcal{U} = (u_1, u_2, \dots, u_m)$ , we define the set of subset sums

$$S(\mathcal{U}) = \left\{ \sum_{j \in J} u_j : J \subseteq [1, m] \right\}.$$

Let  $U = \sum_{j=1}^m u_j$ . Then

$$\{0, U\} \subseteq S(\mathcal{U}) \subseteq [0, U]$$

and  $n \in S(\mathcal{U})$  if and only if  $U - n \in S(\mathcal{U})$ . The sequence  $\mathcal{U}$  is called *complete* if  $S(\mathcal{U}) = [0, U]$ . For example, the sequence  $(1, 2, 3, \dots, m)$  is complete for all  $m \geq 1$ . The sequence  $(1, 1, 3)$  is complete, but the sequence  $(1, 3)$  is not complete. (This is the finite analogue of an infinite complete sequence, which is a sequence  $\mathcal{U}$  of positive integers such that  $S(\mathcal{U})$  contains all sufficiently large integers (cf. Szemerédi-Vu [6]).)

The sequence  $\mathcal{U}$  has *distinct subset sums* if  $|S(\mathcal{U})| = 2^m$ , that is, if the conditions  $I, J \subseteq \{1, 2, \dots, m\}$  and  $\sum_{i \in I} u_i = \sum_{j \in J} u_j$  imply that  $I = J$ . For example, the sequence  $(1, g, g^2, \dots, g^{k-1})$  has distinct subset sums for every  $g \geq 2$ .

If  $\mathcal{U} = (u_1, \dots, u_m)$  is an increasing sequence of positive integers and  $f(x_1, \dots, x_m) = \sum_{j=1}^m u_j x_j$ , then

$$f(1, \dots, 1) = \sum_{j=1}^m u_j = U$$

and

$$f(\{0, 1\}) = S(\mathcal{U}).$$

In particular,  $\mathcal{U}$  is complete if and only if  $f(\{0, 1\}) = [0, U]$ , and  $\mathcal{U}$  has discrete subset sums if and only if  $|f(\{0, 1\})| = 2^m = M_f(2)$ .

For a finite set  $A$  of integers and for integers  $c \neq 0$  and  $d$ , we define the affine transformation

$$c * A + d = \{ca + d : a \in A\}.$$

If  $f(x_1, \dots, x_m) = u_1 x_1 + \dots + u_m x_m \in \mathcal{LF}(m)$  with  $U = u_1 + \dots + u_m$ , then

$$f(c * A + d) = c * f(A) + dU$$

and

$$|f(c * A + d)| = |f(A)|$$

for integers  $c \neq 0$  and  $d$ . Thus, the function  $|f(A)|$  is an affine invariant of  $A$  (cf. Nathanson [4]).

The study of inverse problems for  $m$ -ary forms is related to the paper [5], which initiated the comparative study of binary linear forms.

## 2. A LOWER BOUND FOR $m$ -ARY LINEAR FORMS

The following result is elementary but fundamental.

**Lemma 1.** *Let  $f : \mathbf{Z}^m \rightarrow \mathbf{R}$  be a real-valued function of  $m$  integer variables. Let  $g : \mathbf{Z} \rightarrow \mathbf{R}$  be a strictly increasing function such that*

$$(1) \quad f([a, b]) \subseteq [g(a), g(b)] \quad \text{for all integers } a < b.$$

*Let  $\ell$  and  $\lambda$  be positive integers with  $\ell \geq 2$  such that*

$$(2) \quad N_f(\ell) \geq \lambda.$$

*Let  $A = \{a_i\}_{i=0}^{k-1}$  be a set of  $k$  integers with  $a_{i-1} < a_i$  for  $i = 1, \dots, k-1$ . Let  $k-1 = q(\ell-1) + r$ , where  $0 \leq r \leq \ell-2$ . Define*

$$\mu(A) = |f(\{a_{k-r-1}, a_{k-r}, a_{k-r+1}, \dots, a_{k-1}\})|.$$

*Then*

$$(3) \quad |f(A)| \geq (\lambda - 1) \left( \frac{k - r - 1}{\ell - 1} \right) + \mu(A)$$

*and*

$$(4) \quad N_f(k) \geq \left( \frac{\lambda - 1}{\ell - 1} \right) k - \lambda + 2$$

*for every positive integer  $k$ .*

*Proof.* Dividing  $k$  by  $\ell - 1$ , we have  $k - 1 = q(\ell - 1) + r$ , where  $0 \leq r \leq \ell - 2$ . Then  $k \leq (q + 1)(\ell - 1)$ . Let  $A = \{a_0, a_1, \dots, a_{k-1}\}$  be a set of  $k$  integers, where

$$a_0 < a_1 < \dots < a_{k-1}.$$

For  $j = 0, 1, \dots, q - 1$ , we define the sets

$$A_j = \{a_i : i \in [j(\ell - 1), (j + 1)(\ell - 1)]\}.$$

Let  $A_q$  be the set

$$A_q = \{a_i : i \in [q(\ell - 1), q(\ell - 1) + r]\} = \{a_{k-r-1}, a_{k-r}, a_{k-r+1}, \dots, a_{k-1}\}$$

Then

$$A_{j-1} \cap A_j = \{j(\ell - 1)\}$$

for  $j = 1, \dots, q$ , and

$$A = \bigcup_{j=0}^q A_j.$$

Since

$$\max(A_{j-1}) = a_{j(\ell-1)} = \min(A_j)$$

for  $j = 1, \dots, q$ , and since the function  $g$  is strictly increasing, condition (1) implies that  $f(A_{j'}) \cap f(A_j) = \emptyset$  if  $j - j' > 1$ , and

$$f(A_{j-1}) \cap f(A_j) = \{g(j(\ell - 1))\} = \{f(j(\ell - 1)), \dots, j(\ell - 1)\}$$

for  $j = 1, \dots, q$ . Note that  $\mu(A) = |f(A_q)| = 1$  if  $r = 0$ .

By condition (2), we have  $|f(A_j)| \geq \lambda$  for  $j = 0, 1, \dots, q - 1$ , and so

$$\begin{aligned} |f(A)| &= \left| f\left(\bigcup_{j=0}^q A_j\right) \right| \\ &\geq \sum_{j=0}^q |f(A_j)| - q \\ &\geq \lambda q - q + \mu(A) \\ &= (\lambda - 1) \left( \frac{k - r - 1}{\ell - 1} \right) + \mu(A) \\ &\geq \left( \frac{\lambda - 1}{\ell - 1} \right) k - \lambda + 2. \end{aligned}$$

The observation that the last inequality is independent of the set  $A$  completes the proof.  $\square$

**Lemma 2.** *Let  $f \in \mathcal{LF}(m)$ . If  $\ell$  and  $\lambda$  are positive integers with  $\ell \geq 2$  such that  $N_f(\ell) \geq \lambda$ , then*

$$N_f(k) \geq \left( \frac{\lambda - 1}{\ell - 1} \right) k - \lambda + 2$$

*for every positive integer  $k$ .*

*Proof.* Let  $f(x_1, \dots, x_m) = \sum_{j=1}^m u_j x_j$ . For  $U = \sum_{j=1}^m u_j$ , we define the strictly increasing function  $g(x) = Ux$ . If  $a \leq x_j \leq b$  for  $j = 1, \dots, m$ , then

$$g(a) = Ua \leq f(x_1, \dots, x_m) \leq Ub = g(b)$$

and so

$$f([a, b]) \subseteq [g(a), g(b)]$$

for all integers  $a < b$ . The result follows from Lemma 1.  $\square$

**Theorem 1.** *For all positive integers  $m$  and  $k$ ,*

$$\mathcal{N}_m^*(k) = \left( \frac{m^2 + m}{2} \right) k - \left( \frac{m^2 + m - 2}{2} \right).$$

*Proof.* Let  $f \in \mathcal{LF}^*(m)$ . Then  $f(x_1, x_2, \dots, x_m) = u_1 x_1 + u_2 x_2 + \dots + u_m x_m$  with  $1 \leq u_1 < u_2 < \dots < u_m$ . For integers  $a < b$  and  $i = 0, 1, \dots, m$ , we define the integer

$$\begin{aligned} s_i &= f \left( \underbrace{a, \dots, a}_{m-i \text{ terms}}, \underbrace{b, \dots, b}_i \right) \\ &= (u_1 + \dots + u_{m-i}) a + (u_{m-i+1} + \dots + u_m) b \\ &\in f(A). \end{aligned}$$

Then

$$s_0 < s_1 < \dots < s_m.$$

For  $i = 0, 1, \dots, m-2$  and  $j = 0, 1, \dots, m-i$ , the integer

$$\begin{aligned} t_{i,j} &= f \left( \underbrace{a, \dots, a}_{j-1 \text{ terms}}, b, \underbrace{a, \dots, a}_{m-i-j \text{ terms}}, \underbrace{b, \dots, b}_i \right) \\ &= (u_1 + \dots + u_{j-1}) a + u_j b + (u_{j+1} + \dots + u_{m-i}) a + (u_{m-i+1} + \dots + u_m) b \\ &\in f(A) \end{aligned}$$

satisfies

$$s_i = t_{i,0} < t_{i,1} < \dots < t_{i,m-i-1} < t_{i,m-i} = s_{i+1},$$

It follows that

$$N_f(2) \geq (m+1) + \sum_{i=0}^{m-2} (m-i-1) = \frac{m^2 + m + 2}{2}.$$

Applying Lemma 2 with  $\ell = 2$  and  $\lambda = (m^2 + m + 2)/2$ , we obtain

$$\mathcal{N}_m^*(k) \geq \left( \frac{m^2 + m}{2} \right) k - \left( \frac{m^2 + m - 2}{2} \right).$$

To prove that this lower bound is best possible, we consider the linear form

$$f(x_1, \dots, x_m) = x_1 + 2x_2 + \dots + ix_i + \dots + mx_m \in \mathcal{LF}^*(m)$$

and the finite set

$$A = \{0, 1, \dots, k-1\}.$$

Then

$$f(A) = \left[ 0, \frac{m(m+1)(k-1)}{2} \right]$$

and so

$$|f(A)| = \left( \frac{m^2 + m}{2} \right) k - \left( \frac{m^2 + m - 2}{2} \right) = \mathcal{N}_m^*(k).$$

This completes the proof.  $\square$

## 3. A LOWER BOUND FOR BINARY AND TERNARY LINEAR FORMS

**Theorem 2.** *Let  $f(x_1, x_2) = u_1x_1 + u_2x_2 \in \mathcal{LF}(2)$ , where  $1 \leq u_1 < u_2$  and  $\gcd(u_1, u_2) = 1$ .*

- (i) *If  $f(x_1, x_2) = x_1 + x_2$ , then  $N_f(k) = 2k - 1$ .*
- (ii) *If  $f(x_1, x_2) = x_1 + 2x_2$ , then  $N_f(k) = 3k - 2$ .*
- (iii) *If  $f(x_1, x_2) \neq x_1 + x_2$  or  $x_1 + 2x_2$ , then*

$$N_f(k) \geq \left\lceil \frac{7k - 5}{2} \right\rceil.$$

*Proof.* Let  $|A| = k$ . If  $f(x_1, x_2) = x_1 + x_2$ , then  $|f(A)| \geq 2k - 1$  and  $f([0, k - 1]) = [0, 2k - 2]$ , hence  $|f([0, k - 1])| = 2k - 1$ .

If  $f(x_1, x_2) = x_1 + 2x_2$ , then  $|f(A)| \geq 3k - 2$  by Theorem 1. Moreover,  $f([0, k - 1]) = [0, 3k - 3]$  and so  $|f([0, k - 1])| = 3k - 2$ .

If  $f(x_1, x_2) = u_1x_1 + u_2x_2 \in \mathcal{LF}(2)$  and  $f(x_1, x_2) \neq x_1 + x_2$  or  $x_1 + 2x_2$ , then  $u_2 \geq 3$ . We shall prove that  $N_f(3) = 8$  or  $9$ . We use the fact that the quadratic form  $u_1^2 + u_1u_2 - u_2^2 \neq 0$  for all nonzero integers  $u_1$  and  $u_2$ .

Let  $A = \{a, b, c\}$ , where  $a < b < c$ . Then  $|f(A)| \leq 9$ . We have the following strictly increasing sequence of seven elements of  $f(A)$ :

$$\begin{aligned} u_1a + u_2a &< u_1b + u_2a < u_1a + u_2b < u_1b + u_2b \\ &< u_1c + u_2b < u_1b + u_2c < u_1c + u_2c \end{aligned}$$

and so  $|f(A)| \geq 7$ . There is another strictly increasing sequence of four elements of  $f(A)$ :

$$u_1b + u_2a < u_1c + u_2a < u_1a + u_2c < u_1b + u_2c.$$

If  $|f(A)| = 7$ , then

$$(5) \quad \{u_1c + u_2a, u_1a + u_2c\} \subseteq \{u_1a + u_2b, u_1b + u_2b, u_1c + u_2b\}.$$

This is possible in only three ways. In the first case, we have

$$\begin{aligned} u_1c + u_2a &= u_1a + u_2b \\ u_1a + u_2c &= u_1b + u_2b. \end{aligned}$$

Eliminating  $a$  from these equations, we obtain  $(u_1^2 + u_1u_2 - u_2^2)(c - b) = 0$ , which is false.

In the second case,

$$\begin{aligned} u_1a + u_2c &= u_1b + u_2a \\ u_1c + u_2a &= u_1b + u_2c. \end{aligned}$$

Eliminating  $b$  from these equations, we obtain  $(u_2 - 2u_1)(c - a) = 0$  and so  $2u_1 = u_2$ . Since  $\gcd(u_1, u_2) = 1$ , it follows that  $u_1 = 1$  and  $u_2 = 2$ , which is also false.

In the third case,

$$\begin{aligned} u_1c + u_2a &= u_1b + u_2b \\ u_1a + u_2c &= u_1c + u_2b. \end{aligned}$$

Eliminating  $a$  from these equations, we again obtain  $(u_1^2 + u_1u_2 - u_2^2)(c - b) = 0$ , which is false. It follows that (5) is impossible, and so  $|f(A)| \geq 8$ . Applying Lemma 2 with  $\ell = 3$  and  $\lambda = 8$ , we obtain

$$N_f(k) \geq \frac{7k}{2} - 6.$$

We can improve the constant term by using the more precise inequality (3) in Lemma 1. If  $r = 0$ , then  $\mu(A) = 1$  and

$$|f(A)| \geq 7 \left( \frac{k-1}{2} \right) + 1 = \frac{7k-5}{2}.$$

If  $r = 1$ , then  $\mu(A) = N_f(2) = 4$  and

$$|f(A)| \geq 7 \left( \frac{k-2}{2} \right) + 4 = \frac{7k-6}{2}.$$

This completes the proof.  $\square$

**Lemma 3.** *Let  $f(x_1, x_2, x_3) = u_1x_1 + u_2x_2 + u_3x_3 \in \mathcal{LF}(3)$  with  $1 \leq u_1 \leq u_2 \leq u_3$  and  $\gcd(u_1, u_2, u_3) = 1$ . If  $f \in \mathcal{LF}^*(3)$ , then  $N_f(2) = 7$  or 8, and  $N_f(2) = 8$  if and only if  $u_1 + u_2 \neq u_3$ . Also,  $N_3^*(2) = 7$ .*

*Let  $f \in \mathcal{LF}(3) \setminus \mathcal{LF}^*(3)$ .*

- (i) *If  $u_1 = u_2 = u_3 = 1$ , then  $N_f(2) = 4$ .*
- (ii) *If  $u_1 = u_2$  and  $u_3 = 2u_1$ , then  $N_f(2) = 5$ .*
- (iii) *If  $u_1 = u_2$  and  $u_3 \neq 2u_1$ , then  $N_f(2) = 6$ .*
- (iv) *If  $u_1 < u_2 = u_3$ , then  $N_f(2) = 6$ .*

*Proof.* Let  $f(x_1, x_2, x_3) = u_1x_1 + u_2x_2 + u_3x_3$ , where  $1 \leq u_1 < u_2 < u_3$ . Then

$$\begin{aligned} u_1a + u_2a + u_3a &< u_1b + u_2a + u_3a < u_1a + u_2b + u_3a \\ &< u_1a + u_2a + u_3b < u_1b + u_2a + u_3b \\ &< u_1a + u_2b + u_3b < u_1b + u_2b + u_3b. \end{aligned}$$

These inequalities account for seven of the at most eight elements of the set  $f(A)$ . The remaining element is  $f(a, b, b) = u_1b + u_2b + u_3a$ . Since

$$u_1a + u_2b + u_3a < u_1b + u_2b + u_3a < u_1b + u_2a + u_3b$$

it follows that  $N_f(2) = 7$  if and only if  $u_1a + u_2a + u_3b = u_1b + u_2b + u_3a$ . This is equivalent to  $(u_1 + u_2 - u_3)(b - a) = 0$  or  $u_1 + u_2 = u_3$ . It follows that  $N_3^*(2) = N_f(2) = 7$  if and only if  $u_1 + u_2 = u_3$ .

Identities (i)-(iv) are straightforward calculations.  $\square$

**Theorem 3.** *Let  $f(x_1, x_2, x_3) = u_1x_1 + u_2x_2 + u_3x_3 \in \mathcal{LF}^*(3)$  with  $1 < u_1 < u_2 < u_3$  and  $\gcd(u_1, u_2, u_3) = 1$ . Then  $N_k(f) \geq 6k - 5$ . If  $f \in \mathcal{LF}^*(3)$  and  $u_1 + u_2 \neq u_3$ , then  $N_f(k) \geq 7k - 6$ .*

*Proof.* Applying Theorem 1 with  $m = 3$  gives  $N_k(f) \geq 6k - 5$ . By Lemma 3, if  $f \in \mathcal{LF}^*(3)$  and  $u_1 + u_2 \neq u_3$ , then  $N_f(2) = 8$ . Applying Lemma 2 with  $\ell = 2$  and  $\lambda = 8$  gives  $N_f(k) \geq 7k - 6$ .  $\square$

Note that an increasing sequence  $(u_1, u_2, u_3)$  has distinct subset sums if and only if it is strictly increasing and  $u_1 + u_2 \neq u_3$ .

#### 4. AN INVERSE PROBLEM FOR LINEAR FORMS

Let  $f$  be a linear form in  $m$  variables with positive integral coefficients. The inverse problem for  $f$  is to determine the  $k$ -minimizing sets for  $f$ , that is, to describe the structure of a  $k$ -set  $A$  such that  $|f(A)| = N_f(k)$ . For example, if  $f(x_1, \dots, x_m) = x_1 + \dots + x_m$ , then  $N_f(k) = mk - m + 1$ , and  $N_f(A) = mk - m + 1$  if and only if  $A$  is an arithmetic progression of length  $k$  (Nathanson [3, Theorem 1.6]). If  $f(x_1, x_2) =$

$x_1 + 2x_2$ , then Cilleruelo, Silva, and Vinuesa [1] proved that  $N_f(k) = 3k - 2$ , and  $N_f(A) = 3k - 2$  if and only if  $A$  is an arithmetic progression. This result generalizes to all  $m$ -ary forms whose coefficient sequence is complete.

**Theorem 4.** *Let  $\mathcal{U} = (u_1, \dots, u_m)$  be a complete increasing sequence of positive integers with  $U = \sum_{j=1}^m u_j$ . Consider the linear form  $f(x_1, \dots, x_m) = u_1x_1 + \dots + u_mx_m$ . Then  $N_f(k) = Uk - U + 1$ , and the set  $A$  is a minimizing  $k$ -set for  $f$  if and only if  $A$  is an arithmetic progression of length  $k$ .*

*Proof.* Since  $\mathcal{U}$  is complete, it follows that for any integers  $a$  and  $b$  with  $a < b$  we have

$$\begin{aligned} f(\{a, b\}) &= \left\{ \left( \sum_{i \in [1, m] \setminus I} u_i \right) a + \left( \sum_{i \in I} u_i \right) b : I \subseteq [1, m] \right\} \\ &= \{(U - \ell)a + \ell b : \ell = 0, 1, \dots, U\} \\ &= \{Ua + \ell(b - a) : \ell = 0, 1, \dots, U\}. \end{aligned}$$

Since  $f(\{i - 1, i\}) = [U(i - 1), Ui]$ , it follows that

$$\begin{aligned} [0, U(k - 1)] &= \bigcup_{i=1}^{k-1} [U(i - 1), Ui] = \bigcup_{i=1}^{k-1} f(\{i - 1, i\}) \\ &\subseteq f([0, k - 1]) \subseteq [0, U(k - 1)]. \end{aligned}$$

Then  $f([0, k - 1]) = [0, U(k - 1)]$  and  $N_f(k) \leq |f([0, k - 1])| = Uk - U + 1$ .

Applying Lemma 2 with  $\ell = 2$  and  $\lambda = U + 1$ , we obtain the lower bound  $|f(A)| \geq Uk - U + 1$ , and so  $N_f(k) = Uk - U + 1$ . Since  $|f([0, k - 1])| = Uk - U + 1$  and  $|f(A)|$  is an affine invariant of  $A$ , it follows that  $|f(A)| = Uk - k + 1$  for every arithmetic progression  $A$  of length  $k$ .

Conversely, let  $A = \{a_0, a_1, \dots, a_{k-1}\}$  be a minimizing  $k$ -set for  $f$  with  $a_0 < a_1 < \dots < a_{k-1}$ . Since  $(u_1, \dots, u_m)$  is a complete sequence,

$$f(\{a_{i-1}, a_i\}) = \{(U - \ell)a_{i-1} + \ell a_i : \ell = 0, 1, \dots, U\}.$$

For  $i = 1, \dots, k - 2$  we have the inequalities

$$\begin{aligned} \boxed{(U - 1)a_{i-1} + a_i} &< (U - 2)a_{i-1} + 2a_i < \dots < a_{i-1} + (U - 1)a_i < Ua_i \\ &< (U - 1)a_i + a_{i+1} < \dots < 2a_i + (U - 2)a_{i+1} \\ &< \boxed{a_i + (U - 1)a_{i+1}} < Ua_{i+1}. \end{aligned}$$

Since  $|f(A)| = Uk - U + 1$ , it follows that

$$(6) \quad f(A) = \bigcup_{i=1}^{k-1} f(\{a_{i-1}, a_i\}) = \bigcup_{i=1}^{k-1} \{(U - k)a_{i-1} + ka_i : k = 0, 1, \dots, U\}.$$

We also have

$$\begin{aligned} \boxed{(U - 1)a_{i-1} + a_i} &< (U - 1)a_{i-1} + a_{i+1} < (U - 2)a_{i-1} + 2a_{i+1} < \dots \\ &< 2a_{i-1} + (U - 2)a_{i+1} < a_{i-1} + (U - 1)a_{i+1} \\ &< \boxed{a_i + (U - 1)a_{i+1}}. \end{aligned}$$

Equation (6) implies that

$$(7) \quad \begin{aligned} & \{(U-k)a_{i-1} + ka_{i+1} : k = 1, \dots, U-1\} \\ & \subseteq \{(U-k)a_{i-1} + ka_i : k = 2, \dots, U\} \\ & \cup \{(U-k)a_i + ka_{i+1} : k = 1, \dots, U-2\}. \end{aligned}$$

We want to prove that  $A$  is an arithmetic progression. If not, then  $a_{i-1} + a_{i+1} \neq 2a_i$  for some  $i \in [1, k-2]$ . It follows that for all  $k \in [1, U/2]$  we have

$$(8) \quad (U-k)a_{i-1} + ka_{i+1} \neq (U-2k)a_{i-1} + 2ka_i$$

and

$$(9) \quad ka_{i-1} + (U-k)a_{i+1} \neq 2ka_i + (U-2k)a_{i+1}.$$

Let  $U' = U/2$  if  $U$  is even and  $U' = (U-1)/2$  if  $U$  is odd. Set inclusion (7) implies that

$$(U-1)a_{i-1} + a_{i+1} \geq (U-2)a_{i-1} + 2a_i.$$

Suppose that

$$(U-k)a_{i-1} + ka_{i+1} \geq (U-2k)a_{i-1} + 2ka_i$$

for some  $k \in [1, U'-1]$ . We deduce from inequality (8) that

$$(U-k)a_{i-1} + ka_{i+1} > (U-2k)a_{i-1} + 2ka_i$$

and so, again by (8),

$$(U-k)a_{i-1} + ka_{i+1} \geq (U-2k-1)a_{i-1} + (2k+1)a_i.$$

It follows again from (7) that

$$(U-(k+1))a_{i-1} + (k+1)a_{i+1} \geq (U-2(k+1))a_{i-1} + 2(k+1)a_i.$$

Continuing inductively, we obtain

$$(10) \quad (U-U')a_{i-1} + U'a_{i+1} \geq (U-2U')a_{i-1} + 2U'a_i$$

and so

$$(U-U')a_{i-1} + U'a_{i+1} > (U-2U')a_{i-1} + 2U'a_i.$$

If  $U = 2U'$  is even, this inequality can be rewritten as

$$U'a_{i-1} + U'a_{i+1} \geq Ua_i.$$

If  $U = 2U' + 1$  is odd, inequality (10) becomes

$$(U'+1)a_{i-1} + U'a_{i+1} \geq a_{i-1} + (U-1)a_i.$$

Inequality (8) and set inclusion (7) imply that

$$(U'+1)a_{i-1} + U'a_{i+1} \geq Ua_i.$$

Therefore,

$$U'a_{i-1} + (U'+1)a_{i+1} \geq (U-1)a_i + a_{i+1}.$$

In both cases we have

$$(11) \quad ka_{i-1} + (U-k)a_{i+1} \geq 2ka_i + (U-2k)a_{i+1}$$

for  $k = U'$ .

Suppose that (11) holds for some  $k \in [2, U']$ . Inequality (9) and set inclusion (7) imply that

$$ka_{i-1} + (U-k)a_{i+1} \geq (2k-1)a_i + (U-(2k-1))a_{i+1}.$$

Therefore,

$$(k-1)a_{i-1} + (U - (k-1))a_{i+1} \geq 2(k-1)a_i + (U - 2(k-1))a_{i+1}.$$

Continuing downward inductively, we obtain

$$a_{i-1} + (U-1)a_{i+1} \geq 2a_i + (U-2)a_{i+1}.$$

Since  $a_{i-1} + (U-1)a_{i+1} < a_i + (U-1)a_{i+1}$ , it follows that  $a_{i-1} + (U-1)a_{i+1} = 2a_i + (U-2)a_{i+1}$ , which implies that  $a_{i-1} + a_{i+1} = 2a_i$ . This is a contradiction. Therefore, the minimizing  $k$ -set  $A$  is an arithmetic progression. This completes the proof.  $\square$

## 5. AN UPPER BOUND FOR LINEAR FORMS

We record here some simple estimates for the maximal function  $M_f(k)$ .

**Theorem 5.** *For all  $m$ -ary linear forms  $f \in \mathcal{LF}(m)$  and all positive integers  $k$ ,*

$$(12) \quad k^m \geq M_f(k) \geq \binom{k}{m}.$$

*If  $f \in \mathcal{LF}^*(m)$ , then*

$$(13) \quad M_f(k) \geq k(k-1) \cdots (k-m+1).$$

*If  $U = \{u_1, u_2, \dots, u_m\}$  is an increasing sequence of positive integers with distinct subset sums, and  $f(x_1, \dots, x_m) = u_1x_1 + u_2x_2 + \cdots + u_mx_m \in \mathcal{LF}(m)$ , then*

$$(14) \quad M_f(k) = k^m.$$

*Proof.* Let  $f(x_1, \dots, x_m) = u_1x_1 + u_2x_2 + \cdots + u_mx_m \in \mathcal{LF}(m)$ . The upper bound for  $M_f(k)$  comes from counting the number of  $m$ -tuples of a  $k$ -element set. To obtain the lower bound in (12), choose  $g > mu_m$  and let  $A = \{1, g, g^2, \dots, g^{k-1}\}$ . If  $(r_1, \dots, r_m)$  and  $(s_1, \dots, s_m)$  are  $m$ -tuples of elements of  $[0, k-1]$  such that  $|\{r_1, \dots, r_m\}| = |\{s_1, \dots, s_m\}| = m$  and the  $k$ -sets  $\{r_1, \dots, r_m\} \neq \{s_1, \dots, s_m\}$  are distinct, then the uniqueness of the  $g$ -adic representations of the positive integers implies that  $f(g^{r_1}, \dots, g^{r_m}) \neq f(g^{s_1}, \dots, g^{s_m})$ . This proves (12).

If  $f \in \mathcal{LF}^*(m)$ , then the coefficients  $u_1, \dots, u_m$  are pairwise distinct. If  $(r_1, \dots, r_m)$  and  $(s_1, \dots, s_m)$  are  $m$ -tuples of elements of  $[0, k-1]$  such that  $|\{r_1, \dots, r_m\}| = |\{s_1, \dots, s_m\}| = m$  and the  $m$ -tuples  $(r_1, \dots, r_m)$  and  $(s_1, \dots, s_m)$  are distinct, then the uniqueness of the  $g$ -adic representations of the positive integers implies that  $f(g^{r_1}, \dots, g^{r_m}) \neq f(g^{s_1}, \dots, g^{s_m})$ . This proves (13).

Finally, suppose that the sequence  $(u_1, u_2, \dots, u_m)$  has distinct subset sums. Let  $(r_1, \dots, r_m)$  and  $(s_1, \dots, s_m)$  be  $m$ -tuples of elements of  $[0, k-1]$  such that  $f(g^{r_1}, \dots, g^{r_m}) = f(g^{s_1}, \dots, g^{s_m})$ . For  $d \in [0, k-1]$ , let  $I_d = \{i \in [1, m] : r_i = d\}$  and  $J_d = \{j \in [1, m] : s_j = d\}$ . Then

$$f(g^{r_1}, \dots, g^{r_m}) = \sum_{d=0}^{k-1} \left( \sum_{i \in I_d} u_i \right) g^d$$

and

$$f(g^{s_1}, \dots, g^{s_m}) = \sum_{d=0}^{k-1} \left( \sum_{j \in J_d} u_j \right) g^d.$$

Since

$$\max \left( \sum_{i \in I_d} u_i, \sum_{j \in J_d} u_j \right) \leq mu_m < g$$

it follows that if  $f(g^{r_1}, \dots, g^{r_m}) = f(g^{s_1}, \dots, g^{s_m})$ , then  $I_d = J_d$  for all  $d$  and so  $r_i = s_i$  for  $i = 1, \dots, m$ . Therefore,  $|f(A)| = k^m$ .  $\square$

## 6. OPEN PROBLEMS

- (1) The minimizing  $k$ -sets for linear forms associated to complete sequences are precisely the arithmetic progressions of length  $k$ . Classify all linear forms  $f(x_1, \dots, x_m)$  with the property that the only minimizing  $k$ -sets are arithmetic progressions. In particular, if  $f(x_1, \dots, x_m) = u_1x_1 + \dots + u_mx_m$  is a linear form whose minimizing  $k$ -sets are arithmetic progressions, then is the sequence  $(u_1, \dots, u_m)$  complete?
- (2) Let  $f(x_1, \dots, x_m) = u_1x_1 + \dots + u_mx_m$  be a linear form with  $U = \sum_{j=1}^m u_j$ . Is the sequence  $(u_1, \dots, u_m)$  complete if  $N_f(k) = Uk - U + 1$ ?
- (3) There is no reason to consider only linear forms. Let  $f(x_1, \dots, x_m)$  be a polynomial with integer coefficients. The set  $A$  is a *minimizing  $k$ -set* for  $f$  if  $|f(A)| = N_f(k)$ . Compute  $N_f(k)$  and determine the minimizing  $k$ -sets for  $f$ .
- (4) Let  $s(x_1, x_2) = x_1 + x_2$ . The Freiman philosophy of inverse problems in additive number theory is to deduce structural information about a finite set  $A$  of integers if the sumset  $s(A) = A + A$  is small (cf. Freiman [2]). Analogously, a natural inverse problem for linear forms and, more generally, arbitrary integer-valued polynomials in  $m$  variables, is to deduce information about the finite sets  $A$  of integers such that  $|f(A)| - N_f(A)$  is small.
- (5) For  $f \in \mathcal{LF}(m)$ , define the set

$$\mathcal{E}_f(k) = \{|f(A)| : A \subseteq \mathbf{Z} \text{ and } |A| = k\}.$$

By definition,  $\min(\mathcal{E}_f(k)) = N_f(k)$  and  $\max(\mathcal{E}_f(k)) = M_f(k)$ . For example, if  $f \in \mathcal{LF}(2)$ , then  $\mathcal{E}_f(2) = [3, 4]$ , and, by Lemma 3,  $\mathcal{E}_f(3) = [4, 8]$ . When is the set  $\mathcal{E}_f(k)$  an interval of integers? For every linear form  $f$  and  $e \in \mathcal{E}_f(k)$ , let  $\mathcal{A}_f(e)$  be the set of all  $k$ -sets  $A$  of integers such that  $|f(A)| = e$ . Then  $\{\mathcal{A}_f(e)\}_{e \in \mathcal{E}_f(k)}$  is a partition of the  $k$ -sets of integers. Can one classify the sets in this partition? There are many such questions.

*Acknowledgement.* I wish to thank Manuel Silva for several useful conversations about binary linear forms.

## REFERENCES

- [1] J. Cilleruelo, M. Silva, and C. Vinuesa, personal communication, 2007.
- [2] G. A. Freiman, *Foundations of a structural theory of set addition*, Translations of Mathematical Monographs, vol. 37, American Mathematical Society, Providence, R.I., 1973.
- [3] M. B. Nathanson, *Additive number theory: Inverse problems and the geometry of sumsets*, Graduate Texts in Mathematics, vol. 165, Springer-Verlag, New York, 1996.
- [4] ———, *Affine invariants, relatively prime sets, and a phi function for subsets of  $\{1, 2, \dots, n\}$* , *Integers* **7** (2007), A1: 1–7.
- [5] M. B. Nathanson, K. O’Byrant, B. Orosz, I. Ruzsa, and M. Silva, *Binary linear forms over finite sets of integers*, *Acta Arith.* (2007), to appear.
- [6] E. Szemerédi and V. H. Vu, *Finite and infinite arithmetic progressions in sumsets*, *Ann. of Math.* (2) **163** (2006), no. 1, 1–35.

DEPARTMENT OF MATHEMATICS, LEHMAN COLLEGE (CUNY), BRONX, NEW YORK 10468  
*E-mail address:* `melvyn.nathanson@lehman.cuny.edu`